Integrated Arctic Observation System

Research and Innovation Action under EC Horizon2020

Grant Agreement no. 727890

Project coordinator:

Nansen Environmental and Remote Sensing Center, Norway

Deliverable 5.7

Processing services for iAOS V1

Integration of new processing services

Revised 26 June 2020

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	30 November 2019	Actual submission date:	20 December 2019
Lead beneficiary for preparing the deliverable:		TERRADUE	
Person-months used to produce deliverable:		0.9 pm	

Authors: Hervé Caumont (Terradue), Torill Hamre (NERSC), Espen Storheim (NERSC)

Reviewer: Pedro Goncalves (Terradue)

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
0.1	16/03/2017	Template	Pedro Goncalves
0.2	26/06/2017	Table of Content	Hervé Caumont
0.3	31/10/2019	Consolidated draft for internal review	Hervé Caumont, Torill Hamre
0.4	10/11/2019	Second draft edition for internal review	Hervé Caumont, Torill Hamre
0.5	29/11/2019	Updated draft	Hervé Caumont, Torill Hamre
0.6	20/12/2019	Final Draft	Hervé Caumont, Torill Hamre
1.0	**20/12/2019**	**Approved release for submission**	**Pedro Goncalves**
1.1	25/06/2020	Revised version from EC comments	T.Hamre, E.Storheim, H.Caumont
1.2	26/06/2020	Submission	Kjetil Lygre

Approval **X**	Date: 26 June 2020	Sign. Stein Sandven Stein Sandven

USED PERSON-MONTHS FOR WRITING THIS DELIVERABLE					
No	***Beneficiary***	***PM***	***No***	***Beneficiary***	***PM***
1	**NERSC**	***0,7***	**24**	**TDUE**	***0,2***
2	UiB		**25**	GINR	
3	**IMR**	-	**48**	UNEXE	
4	MISU		**27**	NIVA	
5	**AWI**	-	**28**	CNRS	
6	IOPAN		**29**	U Helsinki	
7	**DTU**	-	**30**	GFZ	
8	AU		**31**	**ARMINES**	-
9	**GEUS**	-	**32**	IGPAN	
10	**FMI**	-	**33**	U SLASKI	
11	UNIS		**34**	BSC	
12	NORDECO		**35**	DNV GL	
13	SMHI		**36**	RIHMI-WDC	
14	USFD		**37**	NIERSC	
15	**NUIM**	-	**38**	WHOI	
16	IFREMER		**39**	SIO	
17	MPG		**40**	UAF	
18	EUROGOOS		**41**	U Laval	
19	EUROCEAN		**42**	ONC	
20	UPM		**43**	NMEFC	
21	UB		**44**	RADI	
22	UHAM		**45**	KOPRI	
23	NORCE		**46**	NIPR	
			47	PRIC	

DISSEMINATION LEVEL		
PU	Public, fully open	**X**
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

EXECUTIVE SUMMARY

The "Integrated Arctic Observation System" (INTAROS) is a 5-year project funded by Horizon 2020 under the Blue Growth Programme. The overall objective of INTAROS is to build an efficient integrated Arctic Observation System (iAOS) by extending, improving and unifying existing systems in the different regions of the Arctic. Within INTAROS, WP5 (Data integration and management) is tasked with designing and implementing evolutions of the cloud platform and tools for the iAOS.

To demonstrate use of the iAOS cloud platform, NERSC has developed two services for extracting sea ice and acoustic parameters from satellite products and in situ observations, respectively.

The source code for both services, along with instructions for installation and use, is made publicly available at:

- https://github.com/ec-intaros/enb-sea-ice-concentration
- https://github.com/ec-intaros/PAMGuide-R-Tutorial

The services can be used as-is, or extended by other scientists and service developers in the Arctic community. Both services will be used in WP6 in the analysis of sea ice and ocean parameters derived from in situ observations and remote sensing data.

The first service integrates time series of sea ice classification (SIC) products generated from remote sensing data (Sentinel-1) to provide sea ice statistics for selected regions. The service is implemented in Jupyter Notebook, extracting a series of CMEMS sea ice concentration products for a given area and time period. Based on the extracted SIC products the sea ice statistics service estimates the mean sea ice concentration for the selected area and period. The resulting statistics product is stored in NetCDF format, which can be readily published using an open source data server such as the THREDDS Data Server (TDS). This enables integration of these products in other services running in the iAOS cloud platform as well as in applications under development in WP6.

The second service processes and characterizes passive acoustic data (from WP2 (Exploitation of existing observing systems) and WP3 (Enhancement of multidisciplinary in situ systems)). This service produces spectrograms and noise statistics plots that can be used for analysis in combination with time series of satellite remote sensing derived parameters (e.g. ice concentration, ice thickness) covering the regions of interest. The plots are provided in a standard format (PNG) that can be opened by a wide range of tools across all major operating systems. The passive acoustics service supports several formats for input data (e.g. WAV, NetCDF) and has been tested on datasets from several sources (NERSC TDS, CNRS, PANGAEA).

Table of Contents

1. Purpose of the document

1.1. Project Overview

The INTAROS project develops an integrated Arctic Observation System (iAOS) by extending, improving and unifying existing systems in the different regions of the Arctic. The project has a strong multidisciplinary focus, aggregating tools for data integration from different domains like atmosphere, ocean, cryosphere and terrestrial sciences, provided by institutions in Europe, North America and Asia. Satellite Earth observation (EO) data plays an important role in the project due to the growing availability of EO data for observing the global climate and environment. In situ observing systems are also considered but are more limited due to logistical constraints and cost limitations due to the sparseness of in situ data. INTAROS is assessing the strengths and weaknesses of existing observing systems and contributes with innovative solutions to fill some of the critical gaps in the in situ observing network.

INTAROS also develops a platform, the iAOS cloud platform, to search for and access data from distributed databases providing a common entry point to data originating from a wide range of observation networks, scientific campaigns and satellite missions, as well as new data generated within the project. This platform leverages state-of-the-art cloud computing technologies to facilitate seamless access to multidisciplinary data, scalable allocation of data storage and computing power for big data processing, integration and analysis including geostatistical methods. The usefulness and functionality of the platform for services development is demonstrated in selected applications.

The evolution into a sustainable Arctic observing system requires coordination, mobilization and cooperation between the existing European and international infrastructures (in-situ and remote, including space-based), the modeling communities and relevant stakeholder groups. INTAROS includes the development of community-based observing systems, where local knowledge is merged with scientific data. The integrated Arctic Observation System aims at enabling better-informed decisions and better-documented processes within key sectors (e.g. local communities, shipping, tourism, fisheries), in order to strengthen the societal and economic role of the Arctic region and support the EU strategy for the Arctic and related maritime and environmental policies.

1.2. Purpose and scope

As shown on Figure 1, INTAROS is organized in eight work packages (WPs) where WP5 activities have connections with other WPs. It has connections with WP2 that assesses existing observing systems and components and prepares their integration into the iAOS platform. From WP3, it integrates new temporal and geographic coverage of in situ observation data in selected regions of the Arctic in order to fill selected gaps in the multidisciplinary observing system. From WP4, it addresses inclusion into the iAOS cloud platform of community-based observing programs with a particular focus on the engagement of the local communities to participate in the development of cloud platform so that they can take full benefit of such a system. In addition, it has connections with WP6 that demonstrates applications to selected stakeholders from governmental institutions, international agencies, industry, local and research communities.

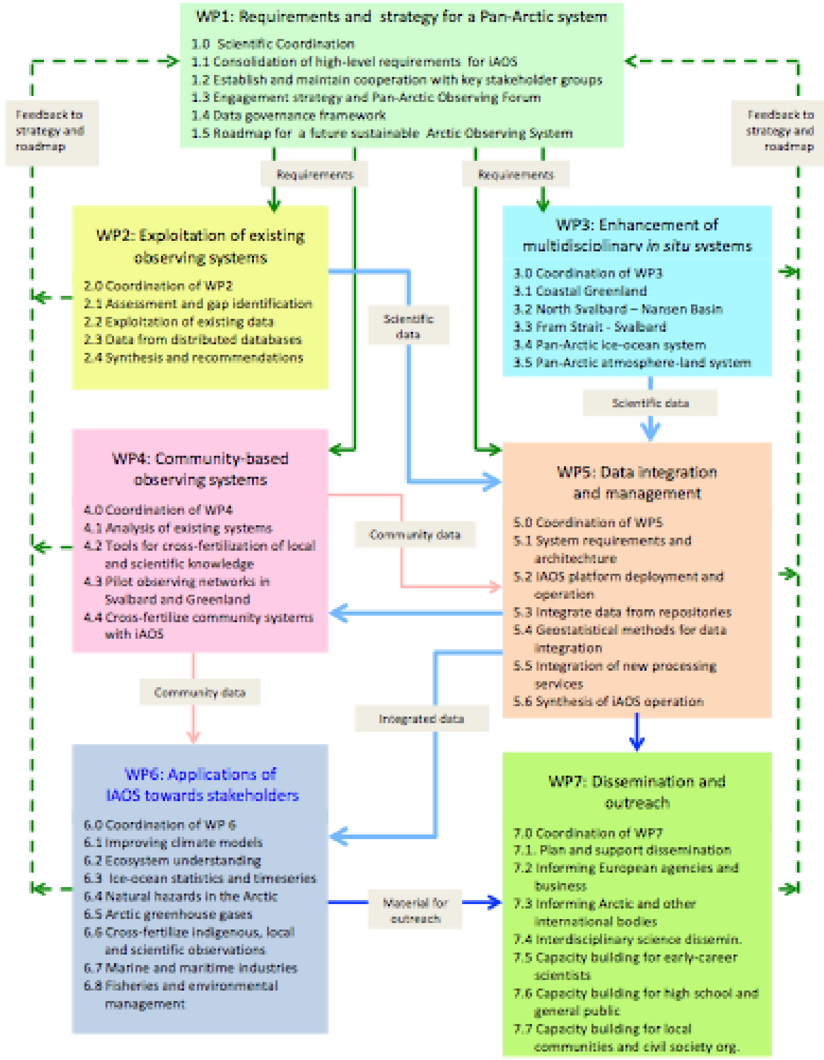


Figure 1. Overall structure of the INTAROS work packages

Under WP5, the new data generated in INTAROS is ingested into existing data repositories to be accessible from the iAOS platform. WP5 has the objective to integrate multidisciplinary and distributed data repositories and to provide a set of tools for data analysis, transformation and visualization. It offers seamless access to observations and derived parameters by allowing the integration of geo-statistical methods for interpolation of spatiotemporal datasets including the new observations from WP2-4, and store the generated datasets in an iAOS enabled repository. The platform facilitates seamless access to multidisciplinary data, scalable allocation of data storage and computer power for integrative data processing and analysis.

This report describes two iAOS services developed by NERSC in WP5. Selected applications from WP6 will further demonstrate the usefulness and functionality of the iAOS cloud platform in service development.

Some of these services that support WP6 are described in Deliverable D5.5 (iAOS platform requirements and architecture consolidation v2).

1.3. Document structure

The structure of the document is as follows:

- Chapter 2 introduces the purpose of the two processing services
- Chapter 3 presents in detail the TS-SIC Data Processing Service Specification
- Chapter 4 presents in detail the C-PAD Data Processing Service Specification
- Chapter 5 provides the conclusions of this work
- Chapter 6 lists the reference publications defining the data processing operations

1.4. Acronyms

CMEMS	Copernicus Marine Environment Monitoring Service
C-PAD	Characterization of Passive Acoustic Data
CEFAS	Centre for Environment, Fisheries & Aquaculture Science
DFT	Discrete Fourier Transform
DOI	Digital Object Identifier
EC	European Commission
EO	Earth Observation
GPL	General Public Licence
iAOS	Integrated Arctic Observation System
INTAROS	Integrated Arctic Observation System
ncdf / NetCDF	Network Common Data Form
NERSC	Nansen Environmental and Remote Sensing Center

OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
PSD	Power Spectrum Density
PSF	Python Software Foundation
SAR	Synthetic Aperture Radar
SIC	Sea Ice Concentration
TOL	Tolerance
TS-SIC	Time Series of Sea Ice Concentration
WIFAR	Waves-in-Ice Forecasting for Arctic Operators
WPS	Web Processing Service (OGC)

2. Introduction

The INTAROS development and implementation of the iAOS cloud platform allows partners to perform the integration and analysis of multidisciplinary data in distributed repositories. The platform is user driven and includes tools for data discovery, aggregation, analysis and visualization. The Ellip Solutions provide an application integration environment to the iAOS software developers who need to build scalable, interoperable data processing applications, and make these applications available as-a-service. Such integration work of data processing services, designed to cover the needs of the Pan-Arctic observing system, is demonstrated via two selected data processing services, made available for use in WP6.

The first service utilises available sea ice concentration products derived from remote sensing data (Sentinel-1) to provide statistics for ice classification in a selected region. A service producing sea ice classification maps from Sentinel-1 SAR has been developed by NERSC in the H2020 NextGEOSS project, which uses the same cloud platform for development as INTAROS. This service is integrated using the Ellip Workflows solution, dedicated to the design of scalable data processing chains. Software developers can access online tools (a Virtual Machine and application framework software) to integrate and test their data processing application, as a highly scalable data processing framework (Hadoop streaming), accessing distributed data sources via a standard interface (OGC OpenSearch) and exposed to client applications also via a standard interface (OGC WPS). In INTAROS WP5, NERSC has developed a sea ice statistics service that will draw upon the results from the sea ice classification service as well as other sea ice concentration products developed by e.g. CMEMS. A first version of this service is described in this report. The service is developed using Jupyter Notebook, which is also supported by the Ellip Workflows solution.

The second service processes and characterizes passive acoustic data (from WP2, WP3 and external sources), for analysis in combination with time series of satellite remote sensing derived parameters (e.g. ice concentration, ice thickness) covering the regions of interest. It is integrated using the Ellip Notebooks solution, dedicated to the creation of interactive laboratory notebooks. Software developers using the iAOS cloud development platform can access online tools (a JupyterLab environment) to integrate and test their data processing functions, as a reproducible experiment.

2.1. iAOS Service “Time Series of Sea Ice Concentration (TS-SIC)”

Service purpose:

This service will enable users to extract a time series of sea ice concentration maps for a selected time period in the Fram Strait and North of Svalbard. The data in the time series can be used to generate different statistics for sea ice concentration in the region.

Products delivered:

Maps of monthly sea ice statistics (e.g. 10% ice, 50% ice, 70% ice, 90% ice) for the selected areas.

2.2. iAOS Service "Characterization of Passive Acoustic Data (C-PAD)"

Service purpose:

This service will enable users to analyse and present passive acoustic data, to support detection and classification of different sources of noise pollution in the ocean.

Products delivered:

Noise level statistics and spectrograms.

3. TS-SIC Service Specification

The iAOS Service "Time Series of Sea Ice Concentration (TS-SIC)" will enable users to extract a time series of sea ice classification maps for a selected time period in the Fram Strait and North of Svalbard. The data in the time series can be used to generate different statistics for sea ice concentration in the region. The products to be delivered on iAOS will address maps of monthly sea ice concentration for the selected areas.

3.1. User Stories

The user stories presented hereafter are simple descriptions of a feature, presented from the perspective of the person who desires such capability as a user or customer of the system.

3.1.1. Roles definition

The following roles are defined for the TS-SIC Service:

Scientist - sea ice scientist working in the INTAROS project.

Decision-maker (offshore operations) - user of service, to be invoked through WP6 (Task 6.3).

Stakeholder - member of the wider sea ice science and offshore operations communities.

3.1.2. User required functionalities

1. Story - Monthly sea ice statistics from Sentinel-1 SAR

Story: As a scientist, I would like to generate sea ice statistics for certain months of the year, to support risk assessment for offshore operators in Arctic waters.

Story exemplified for scientist John:

1. John selects the service for sea ice statistics
2. John defines the area of interest, and the desired start and end date
3. iAOS searches for Sentinel-1 SAR images within the given area and time range
4. iAOS presents list of the images that satisfy the search criteria
5. John selects which of the available images to classify
6. John gives input for the statistics (point-radius or transect, time range, ...)
7. iAOS runs the sea ice statistics service for the selected SAR images
8. iAOS presents the resulting sea ice statistics
9. John selects to store the results through iAOS, and gets a DOI for the new dataset (optional)
10. John selects to download the results (optional)

2. Story - Monthly sea ice statistics from CMEMS sea ice concentration product

Story: As a scientist, I would like to generate sea ice statistics for certain months of the year, to support risk assessment for offshore operators in Arctic waters.

Story exemplified for scientist John:

1. John selects the service for sea ice statistics
2. John defines the area of interest, and the desired start and end date
3. iAOS searches for CMEMS sea ice concentration products within the given area and time range
4. iAOS presents list of the CMEMS SIC products that satisfy the search criteria
6. John gives input for the statistics (point-radius or transect, time range, ...)
7. iAOS runs the sea ice statistics service for the selected CMEMS SIC maps

8. iAOS presents the resulting sea ice statistics
9. John selects to store the results through iAOS, and gets a DOI for the new dataset (optional)
10. John selects to download the results (optional)

User Stories are centered on expected results and benefits from the perspective of the system actors and their roles in operating the system. From there, Use Cases can be more granular, and describe how the system will act to match the expectations.

3.2. Use Cases

Use Case specifications are meant to capture user (actor) point of view while describing functional requirements of the system. They describe the step by step process a user goes through to complete that goal using a software system.

3.2.1. Actors definition

The following actors have been defined for this service:

Scientist - ocean acoustics scientist working in the INTAROS project.

Decision-maker (offshore operations) - user of service, to be invoked through WP6 (Task 6.3).

Stakeholder - member of sea ice science, transport and offshore operations sector.

3.2.2. Actors interactions with the Service

Use Case ID: TSSIC-1

Use Case Name: **Generate Monthly Sea Ice Statistics**

Primary Actor: Scientist

Stakeholders and Interests: Decision-maker, Stakeholder

Preconditions: Scientist is identified and authenticated. Scientist is a registered user of Zenodo (https://zenodo.org/).

Postconditions: Average sea ice concentration is computed for a user defined area and time period.

Main success scenario:

1. Scientist selects a geographic area and a time range of 1 month

2. Service downloads sea ice concentration data for given area and time range
3. Service calculates average sea ice concentration for this area and period
4. Service displays the computed sea ice statistics
5. Scientist selects to store the sea ice statistics through iAOS
6. Scientist selects Zenodo as data centre and provides login information
7. Service generates metadata to accompany the sea ice statistics
8. Service forwards the sea ice statistics and metadata Zenodo, using the provided login information of Scientist
9. Service displays the DOI for the newly registered dataset of sea ice statistics

3.3 Service implementation and examples of use

NERSC has implemented the TS-SIC service in Python. TS-SIC draws upon open source libraries for reading and manipulating scientific data in NetCDF files (as described in Section 3.4). The choice of open source software enables other scientists and service developers to quickly put the service into use, without committing resources to commercial tools.

3.3.1 Main features of the TS-SIC Service

The TS-SIC service generates sea ice concentration statistics for a user defined geographic area and time period. The main features of the current version of the service include:

- TS-SIC utilises the CMEMS product "Svalbard ice chart" provided by the Meteorological Institute of Norway, which is a 1km daily product with sea ice concentration (%) for the Svalbard and Barents Sea region. This CMEMS product is available from 4 January 2010 to present.
- The area supported by TS-SIC is limited to 80°W-80°E ; 60°N-85°N.
- The sea ice concentration statistics is available on a monthly basis.
- The sea concentration ice statistics is provided as a NetCDF file.
- A simple plot of the monthly mean sea ice concentration is provided in a PNG file.

3.3.2 Examples of sea ice statistics products generated using TS-SIC

The TS-SIC service was used to generate monthly mean sea ice concentration (SIC) fields using the CMEMS daily SIC product for the Svalbard region. Figures 2 and 3 show the mean SIC fields for June - September 2018 and June - September 2019, respectively. The plots illustrate how the sea ice conditions can vary significantly from one year to the next. In summer of 2018, large areas north of Svalbard and towards Franz Josef Land were ice free already in June. While in 2019 the ice conditions were much more severe in this region, with heavy ice throughout the summer.

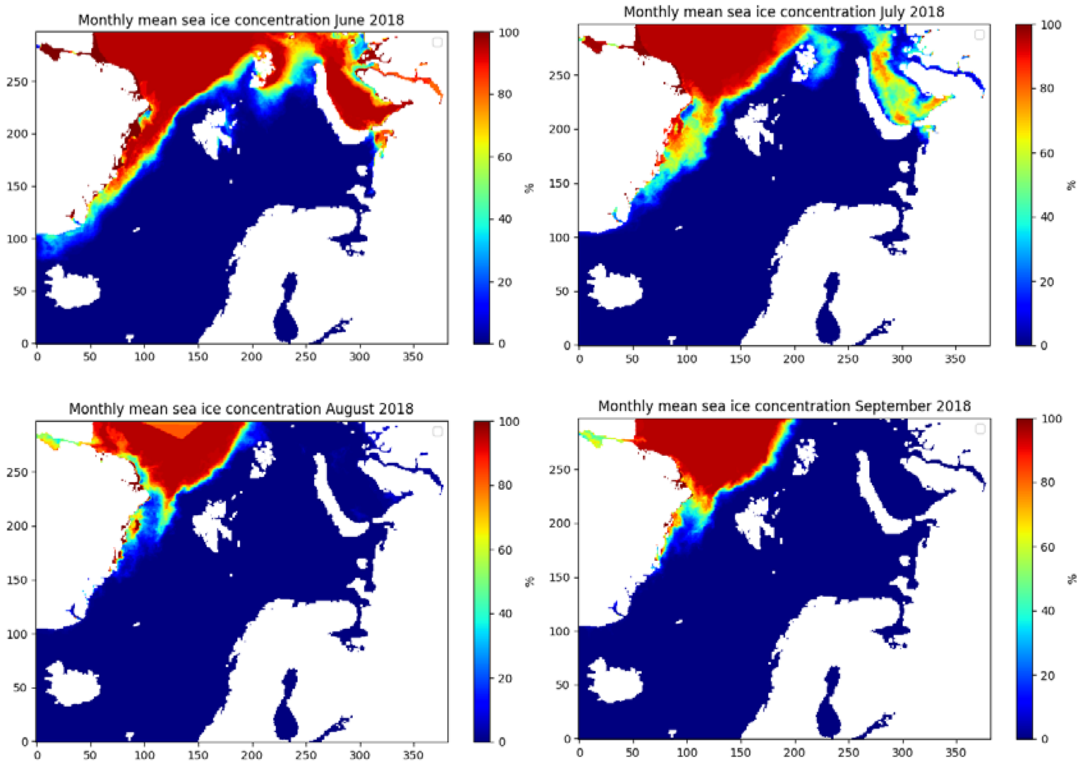
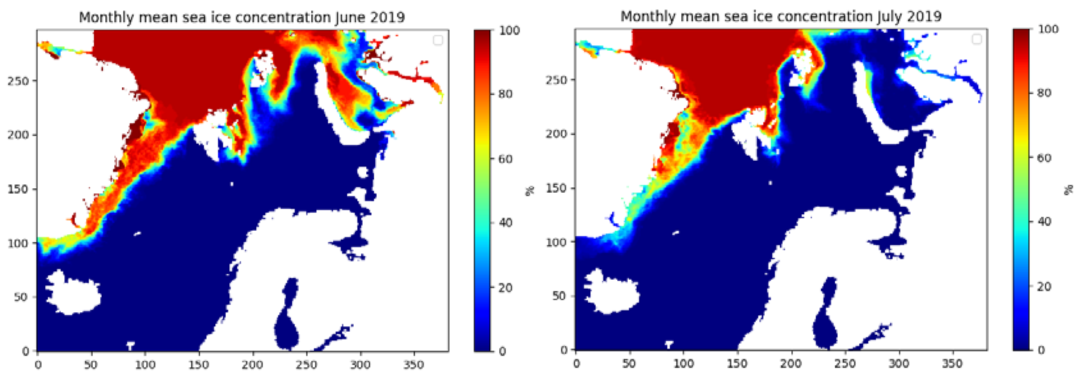


Figure 2. Examples of sea ice concentration statistics, monthly means for June - September 2018 in the Fram Strait of area north of Svalbard, generated from daily CMEMS products.



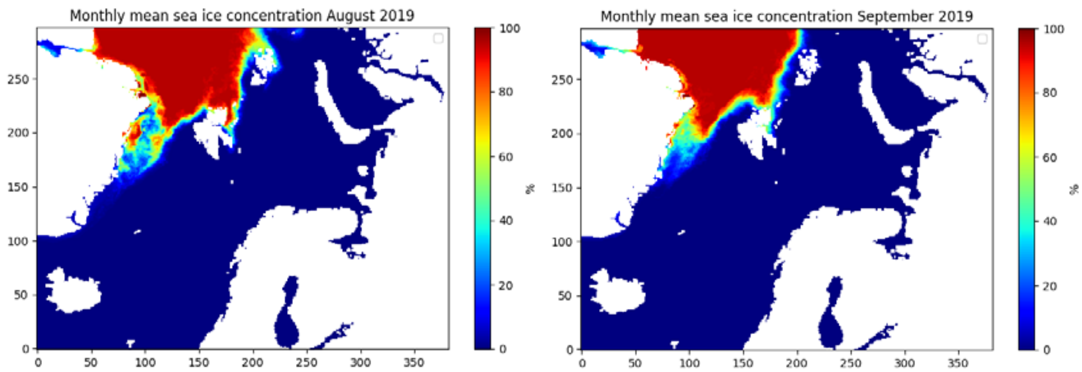


Figure 3. Examples of sea ice concentration statistics, monthly means for June - September 2019 in the Fram Strait of area north of Svalbard, generated from daily CMEMS products.

3.3.3 How to access the TS-SIC Service

The TS-SIC iAOS Service can be used by partners in WP6 to generate mean sea ice concentration for user defined periods between January 2010 and present. Partners can modify the service to generate statistics for longer periods than one month or for seasonal means for the last decade. Source code for the service and the software it depends on is publicly available (Table 1).

For use in WP6, the TS-SIC iAOS Service will be made accessible through the iAOS Portal.

3.3.4 IPR and licensing summary table for TS-SIC Service

The software package for the TS-SIC iAOS Service is available from a public git repository (Table 1). Besides source code, this repository provides a description of dependencies and installation instructions.

Table 1. Online repository for the TS-SIC iAOS Service.

Package	Source URL	License
TS-SIC	https://github.com/ec-intaros/enb-sea-ice-concentration	CC BY 4.0

3.3.5 Roadmap for future development

The TS-SIC Service will be further developed in collaboration with WP6, in particular with partners in Task 6.3 ("Ice-ocean statistics for decisions support and risk assessment").

The service will be used to mean monthly means of sea ice concentration for inclusion in data analysis in WP6. A possible extension to the service is to integrate SIC products from other data providers to be able to extend the area or time range of the generated statistics. Products generated will be made available through the INTAROS Data Catalogue.

3.4 Software modules required by the TS-SIC service

The TS-SIC iAOS Service is developed using widely used open source tools for manipulating scientific data in NetCDF files (netcdf4) and for plotting data held in NetCDF (matplotlib).

Both of these tools have extensive online documentation and tutorials, as well as an active community helping programmers in efficiently using these tools. Both support Python 3.

3.4.1 Required *software*

This service has been developed using Python 3 and Jupyter Notebook, using open source tools netcdf4 and matplotlib.

The netcdf4 library is used for reading NetCDF files provided by an OPeDAP server and writing NetCDF files for the generated monthly mean sea ice concentration fields. matplotlib is used for generating and storing plots of the monthly fields in PNG files.

3.4.2 IPR and licensing summary table for required software modules

The following software packages were used to develop the TS-SIC iAOS Service.

Table 2. Online repository for software modules required by the TS-SIC iAOS Service.

Package	**Source URL**	**License**
netcdf4	https://github.com/Unidata/netcdf4-python	Copyright 2008 by Jeffrey Whitaker
matplotlib	https://matplotlib.org/	https://matplotlib.org/users/license.html (based on PSF - Python Software Foundation License)

3.4.3 Online documentation resources

The following resources are recommended for scientists and service developers that want to run the TS-SIC service or customise it for their needs:

- netcdf4 module documentation: https://unidata.github.io/netcdf4-python/netCDF4/index.html
- Open-source Project for a Network Data Access Protocol (OPeNDAP) http://www.opendap.org/
- matplotlib tutorials: https://matplotlib.org/tutorials/index.html
- CMEMS PRODUCT USER MANUAL (CMEMS-SI-PUM-011-002) https://marine.copernicus.eu/documents/PUM/CMEMS-SI-PUM-011-002.pdf
- CMEMS QUALITY INFORMATION DOCUMENT (CMEMS-SI-QUID-011-001TO007-009TO014) https://marine.copernicus.eu/documents/QUID/CMEMS-SI-QUID-011-001to007-009to014.pdf

3.5 Cloud resources allocation

3.5.1 Data resources

This service uses Sentinel-1 SAR based sea ice concentration products from CMEMS.

3.5.2 ICT resources in the iAOS cloud platform

This service has been developed using Python and Jupyter Notebook. The next step is to install the notebook and the needed tools in the iAOS cloud platform. This will enable other scientists to reuse the notebook for generation of sea ice concentration statistics for areas and time periods of their interest.

4. C-PAD Service Specification

The iAOS Service "Characterization of Passive Acoustic Data (C-PAD)" will enable users to analyse and present passive acoustic data, to support detection and classification of different sources of noise pollution in the ocean. The products to be delivered on iAOS will address Noise level statistics and spectrograms.

4.1. User Stories

The user stories presented hereafter are simple descriptions of a feature, presented from the perspective of the person who desires such capability as a user or customer of the system.

4.1.1. Roles definition

The following roles are defined for the C-PAD Service:

Scientist - ocean acoustics scientist working in the INTAROS project.

Decision-maker (offshore operations) - user of service, to be invoked through WP6 (Task 6.3).

Stakeholder - member of the ocean acoustics and marine environmental conditions communities.

4.1.2. User required functionalities

1. **Story - Passive acoustic data analysis and visualisation**

Story: As a scientist, I would like to process passive acoustic data to study the ocean soundscape.

Story exemplified for scientist Bob:

1.Bob selects the service for processing acoustic data

2.Bob defines an area and time range of interest

3.iAOS searches for passive acoustic datasets fulfilling these criteria

4.iAOS presents a list of these datasets

5.Bob selects one of the datasets for processing

6.iAOS runs the passive acoustic data processing service for the selected dataset

7.iAOS presents the estimated ocean soundscape parameters

8.Bob selects to download the results (optional)

User Stories are centered on expected results and benefits from the perspective of the system actors and their roles in operating the system. From there, Use Cases can be more granular, and describe how the system will act to match the expectations.

4.2 Use Cases

Use Case specifications are meant to capture user (actor) point of view while describing functional requirements of the system. They describe the step by step process a user goes through to complete that goal using a software system.

4.2.1 Actors definition

The following actors have been defined for this service:

Scientist - ocean acoustics scientist working in the INTAROS project.

Decision-maker (offshore operations) - user of service, to be invoked through WP6 (Task 6.3).

Stakeholder - member of the ocean acoustics and marine environmental communities.

4.2.2 Actors interactions with the service

The use case is based on a simple interaction sequence where the user can read input data from different sources as follows:

1. as a WAV file previously uploaded on the JupyterLab workspace hosted on the Ellip Notebooks solution,
2. as a NetCDF data served from a Thredds server

Then, the user is supported by the application to calibrate the data, and plot the output from the different analysis types available in PAMGuide. The service is implemented as a Jupyter Notebook, and is made available from an online git repository.

Use Case ID: CPAD-1

Use Case Name: **Analyse Passive Acoustic Data**

Primary Actor: Scientist

Stakeholders and Interests: Decision-maker, Stakeholder

Preconditions: Scientist is identified and authenticated.

Postconditions: Passive acoustic data is calibrated correctly. Power spectrum is computed.

Main success scenario:

1. Scientist selects a file with passive acoustics data
2. Service opens the file and reads the metadata
3. Service displays the metadata

4. Scientist identifies and extracts the calibration metadata
5. Scientist enters calibration metadata as input to the Service
6. Service performs data calibration
7. Scientist selects DFT analysis mode
8. Service performs discrete fourier transform analysis
9. Scientist selects power spectrum density (PSD) spectrogram as plot type
10. Service generates and displays a PSD spectrogram

4.3 Service implementation and examples of use

NERSC has implemented the C-PAD service in R, utilizing Jupyter Notebook as a rapid development and interactive testing tool. C-PAD draws upon several open source tools and libraries for acoustic data processing, analysis and visualisation (as described in Section 4.4). This enables other scientists and service developers to quickly put the service into use, without committing resources to commercial tools.

The C-PAD service is based on the PAMGuide software (Merchant *et al.*, 2015), a tool for analysis of passive acoustic data. This tool was initially designed to process audio files (WAV files), but has now been augmented to support vector input. This allows for processing of data stored in e.g. NetCDF format without converting to WAV before processing, which potentially could lead to loss of data, as well as adding computational time.

4.3.1 Main features of the C-PAD Service

The C-PAD service can read passive acoustic data from either a WAV file or a NetCDF file, perform calibration and processing of these data, and present the results in a spectrogram and a noise statistics plot. The calibration factors must be extracted from the input files or external sources (e.g. cruise report or scientific papers), and entered into the service. If the calibration factors are not known, the service will still run, but users must take into account that the lack of calibration increases the uncertainty of the results.

A spectrogram visualises temporal variation of the spectrum of frequencies in the acoustic signal. This enables the user to visually analyse the changes in different frequency bands over time and possibly detect events such as earthquakes, ship traffic and seismic airgun shots. The analysis of acoustic data can also lead to identification of marine mammals through matching the signal variations of the acoustic recordings to known vocalisations of a given species. The noise statistics generated by the C-PAD service visualises the power spectrum (intensity) of the signal for the different frequencies which the signal is composed of. A series of percentiles (numbers of recordings below a certain intensity) is used to indicate how the signal intensity varies with frequency. A “spike” in this diagram indicates that signal strength is close to the same level throughout the period for which the statistics is computed. Whereas a larger difference between high and low percentiles indicates that the signal varies in strength for a particular frequency for the time period of the statistics.

4.3.2 Examples of passive acoustics data analysis and visualisation using C-PAD

This section shows the C-PAD service applied to passive acoustic data from three sources:

- WIFAR: Waves-in-Ice Forecasting for Arctic Operators; data collected by NERSC; https://www.nersc.no/project/wifar
- INTAROS: Integrated Arctic Observation System; data collected by CNRS; http://www.intaros.eu/
- PerenniAL Acoustic Observatory in the Antarctic Ocean (PALAOA); data collected by AWI; https://doi.pangaea.de/10.1594/PANGAEA.773610

The WIFAR project conducted several field experiments in the Fram Strait and Barents Sea Marginal Ice Zone (MIZ) using ice-going vessels. As part of WIFAR, an integrated ice station was deployed in the Fram Strait during August 2012 and September 2013, for both periods collecting acoustic data from a hydrophone mounted under the ice. Figure 4 shows the spectrogram and noise statistics for an acoustic recording collected during the WIFAR project in August 2012 (Geyer *et al*, 2017;https://doi.org/doi:10.11582/2017.00012). The spectrogram shows a strong, constant signal around 20 Hz throughout the 10 minute recording. This frequency is typical for fin whale vocalization. A peak is also seen near 300 Hz which could be ship noise. This dataset is publicly available through the Norwegian Marine Data Centre (NMDC).

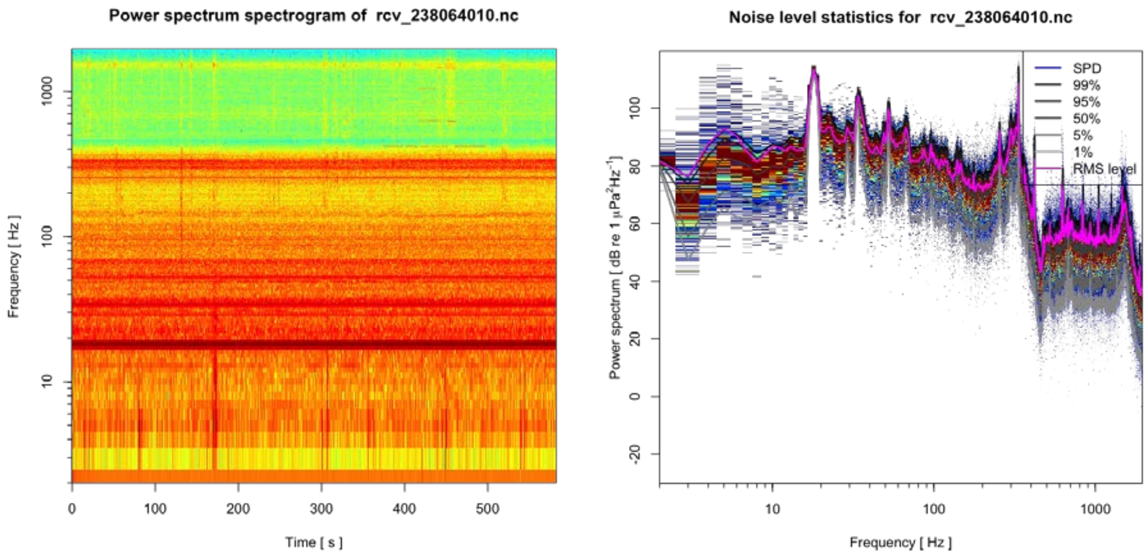


Figure 4. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing WIFAR data (Geyer *et al.*, 2016).

In the INTAROS project, the National Center for Scientific Research (CNRS), France has collected passive acoustic data from a mooring in Kongsfjorden, Svalbard. Figure 5 shows the spectrogram and noise statistics for an acoustic recording collected during the INTAROS project in September 2018 (Chauvaud *et al.*, 2020). The spectrum is dominated by low-frequency noise below 10 Hz. Local peaks around 10 Hz and 80 Hz are also seen, which could be mammal vocalization. Intermittent broad-band signals are also seen in the spectrogram. This dataset is publicly available through the INTAROS Data Catalog (https://catalog-intaros.nersc.no/dataset/passive-acoustic-data-from-kongsfjorden-svalbard-september-2018).

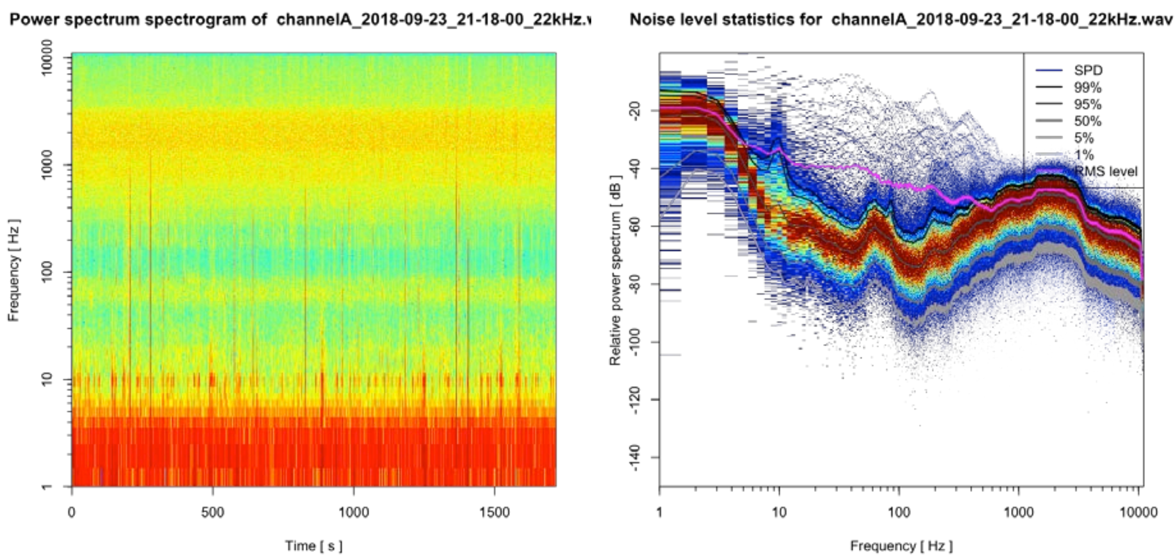


Figure 5. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing acoustic data collected in Kongsfjorden, Svalbard, during the INTAROS project.

The PerenniAL Acoustic Observatory in the Antarctic Ocean (PALAOA) has collected acoustic data off the Ekström ice shelf, Antarctica, for many years since its establishment in December 2005. Recordings from PALAOA are publicly available in PANGAEA, in WAV format. Figure 6 shows the spectrogram and noise statistics for an acoustic recording collected at PALAOA (Kindermann, 2013). This recording is very stable throughout the 10 minute recording period. Peaks are seen from 40 Hz to 400 Hz, but are difficult to identify without further analysis.

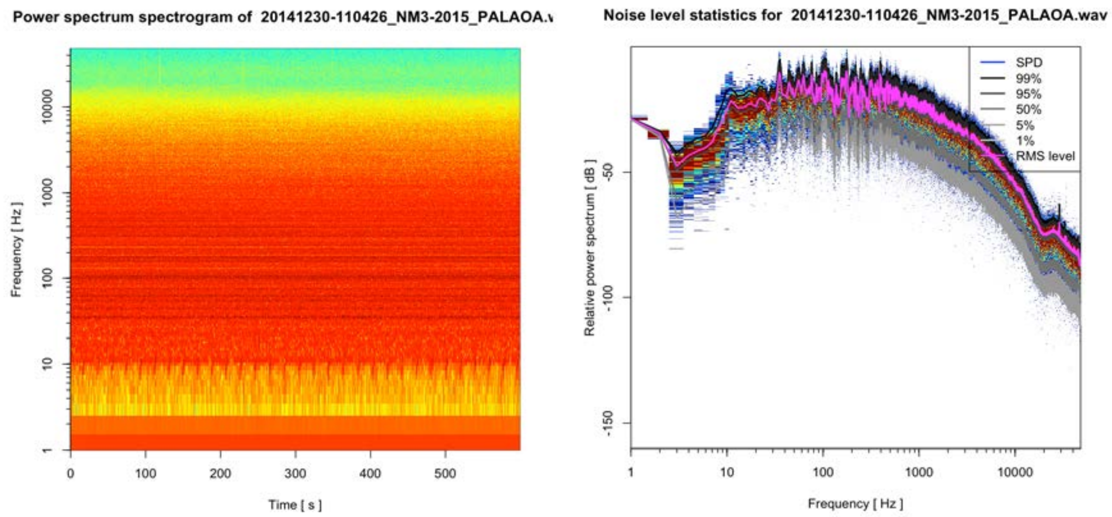


Figure 6. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing passive acoustic data from the PALAOA observatory in Antarctica (https://doi.pangaea.de/10.1594/PANGAEA.773610).

4.3.3 How to access the C-PAD Service

The C-PAD iAOS Service can be used as a basis for further development by scientists and service developers in WP6 and by external parties. Source code for the service and the software it depends on is publicly available (Table 3).

For use in WP6, the C-PAD iAOS Service will be made accessible through the iAOS Portal.

4.3.4 IPR and licensing summary table for C-PAD Service

The software package for the C-PAD Service is available from a public git repository (Table 3). Besides source code, this repository provides a description of dependencies and installation instructions, as well as links to some examples dataset that can be used to test the service.

Table 3. Online repository for the C-PAD iAOS Service.

Package	Source URL	License
C-PAD	https://github.com/ec-intaros/PAMGuide-R-Tutorial	CC BY 4.0

4.3.5 Roadmap for future development

The C-PAD Service will be further developed in collaboration with partners in Task 6.3 ("Ice-ocean statistics for decisions support and risk assessment"). The service will be used to analyse additional datasets depending on the needs emerging from this task. A possible extension to the service is to store the data underlying the spectrogram and noise statistics plots in NetCDF files. This will enable applications under development in WP6 to integrate these data and combine with other data from INTAROS as well as other projects.

4.4 Software modules required by the C-PAD service

The C-PAD Service is developed using PAMGuide, an open source tool for processing and visualisation of passive acoustic data. PAMGuide is written by Nathan Merchant, Centre for Environment, Fisheries & Aquaculture Science (Cefas), United Kingdom. The software package is available for both MATLAB and R development of passive acoustic data services. A detailed description of how to use PAMGuide is given in the supplementary material of Merchant *et al.* (2015).

4.4.1 Required *software*

This service has been developed using the ***R version of PAMGuide***. This software package can read WAV files with passive acoustic data, analyse and visualise these data. Supported plots include spectrograms and noise statistics. Figure 7 illustrates a typical use of PAMGuide. After reading acoustic data into the system, data can be calibrated using information in the metadata or provided in accompanying documentation such as cruise reports. If calibration information is not available, the acoustic data can still be processed and analysed, but the user must take into account that the underlying data is uncalibrated and thus have a larger uncertainty. PAMGuide supports several types of plots, of which spectrograms are the most commonly used to visualise temporal variation of the signal. For statistics analysis, PAMGuide can present the full specter percentiles or the ⅓ octave band level percentiles to enable users to analyse statistics of the acoustic recordings.

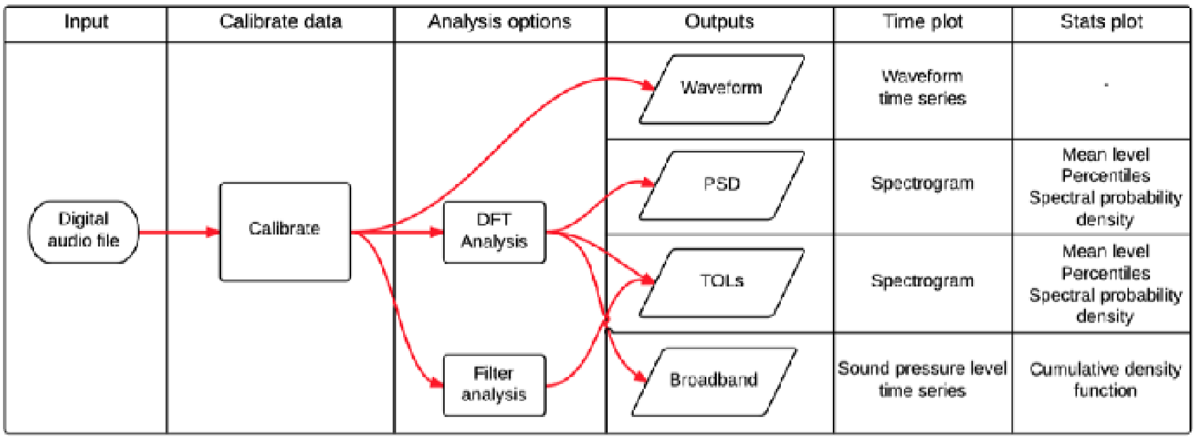


Figure 7. Data flow in PAMGuide (from supplement to Merchant *et al.* (2015)). DFT = discrete Fourier transform; PSD = power spectral density; TOL = 1/3-octave band level.

The R version of PAMGuide uses the ***tuneR package*** to read .WAV files and analyse the data.

In addition, the Characterization of Passive Acoustic Data (C-PAD) Service uses the ***ncdf4 package*** to read passive acoustic data from NetCDF files. This extends PAMGuide to support this widely used data format

4.4.2 IPR and licensing summary table for required software modules

The following software packages were used to develop the C-PAD iAOS Service.

Table 4. Online repository for software modules required by the C-PAD iAOS Service.

Package	Source URL	License
PAMGuide	https://sourceforge.net/projects/pamguide/	Copyright (c) 2014 The Authors.
tuneR	https://cran.r-project.org/package=tuneR	GPL-2 | GPL-3
ncdf4	https://cran.r-project.org/package=ncdf4	GPL (≥ 3)

4.4.3 Online documentation resources

The following resources are recommended for scientists and service developers that want to run the C-PAD service or customise it for their needs:

- PAMGuide Tutorial:
  http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12330/suppinfo
- tuneR Reference Manual:
  https://cran.r-project.org/web/packages/tuneR/tuneR.pdf
- ncdf4 Reference Manual:
  https://cran.r-project.org/web/packages/ncdf4/ncdf4.pdf

4.5 Cloud resources allocation

4.5.1 Data resources

This service does not use any satellite EO data.

This service analyses and visualises passive acoustic data. Sample passive acoustic data from the WIFAR (Waves-in-Ice Forecasting for Arctic Operators) project are provided by the Nansen Center (NERSC), and sample acoustic data from INTAROS is provided by CNRS.

4.5.2 ICT resources in the iAOS cloud platform

This service has been developed using Jupyter Notebook, and a first beta version has been successfully tested in the Sandbox of the iAOS cloud platform. The next step is to install PAMGuide and needed libraries in the iAOS cloud platform. This will enable other scientists to reuse that notebook for analysis of passive acoustic data in WAV or NetCDF format.

5. Conclusion

A first version of each iAOS data processing service "Time Series of Sea Ice Concentration (TS-SIC)" and "Characterization of Passive Acoustic Data (C-PAD)" have been implemented by NERSC using the iAOS service integration tools "Ellip Solutions" provided by Terradue.

These services are available to the iAOS partners for review and reuse based on need, as well as for interacting with NERSC to suggest any improvements for future versions.

The TS-SIC service enables users to extract a time series of sea ice concentration maps for a selected time period in the Fram Strait and North of Svalbard. Based on this time series of daily sea ice concentration fields, a monthly mean sea ice concentration is computed.

The C-PAD service enables users to analyse and present passive acoustic data, to support detection and classification of different sources of noise pollution in the ocean.

6. References

Chauvaud, L., Mathias, D., Jolivet, A., Amice, E., Le Bec, T., Marec, C., Merny, P., and collaborators (2020). Underwater soundscape in Kongsfjorden. Presentation at online INTAROS Ocean Workshop – April 2020.

Geyer, F., Sagen, H., and P. F. Worcester (2017). "WIFAR/UNDER-ICE acoustic recording from an integrated ice station in the Fram Strait marginal ice zone in 2012". NIRD Research Data Archive, https://doi.org/doi:10.11582/2017.00012

Geyer, F., Sagen, H., Hope, G., Babiker, M., and P. F. Worcester, " Identification and quantification of soundscape components in the Marginal Ice Zone," J. Acoust. Soc. Am. 139, 1873–1885 (2016). https://doi.org/10.1121/1.4945989

Kindermann, Lars (2013): Acoustic records of the underwater soundscape at PALAOA with links to audio stream files, 2005-2011. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA, https://doi.org/10.1594/PANGAEA.773610

Merchant, N. D., Fristrup, K. M., Johnson, M. P., Tyack, P. L., Witt, M. J., Blondel, P. and Parks, S. E. (2015), Measuring acoustic habitats. Methods Ecol Evol, 6: 257-265. doi:10.1111/2041-210X.12330

----------- END of DOCUMENT-----------

This report is made under the project

Integrated Arctic Observation System (INTAROS)

funded by the European Commission Horizon 2020 program

Grant Agreement no. 727890.

Project partners: